DOCUMENT RESUME

ED 414 320 TM 027 838

AUTHOR Auchter, Joan E.; Stansfield, Charles W.
TITLE Developing Parallel Tests across Languages: Focus on the Translation and Adaptation Process.
PUB DATE 1997-00-00
NOTE 25p.; Version of a paper presented at the Annual Large Scale Assessment Conference (27th, Colorado Springs, CO, June 15-18, 1997).
PUB TYPE Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Adult Education; Bilingual Students; Educational Attainment; *Educational Certificates; *Equivalency Tests; *High School Equivalency Programs; Scores; *Spanish; *Test Construction; Test Format; Test Items; Test Validity; *Translation
IDENTIFIERS *GED Writing Skills Test; Puerto Rico

ABSTRACT

This paper describes the General Educational Development (GED) Testing Service's Spanish Test Adaptation Project. The GED Tests are designed to give those who have not graduated from high school the opportunity to earn a diploma that is recognized by institutions of higher education and employers. The purpose of this project is to develop, based on an analysis of the issues, Spanish-language versions of the GED tests that parallel the English-language versions so that the GED candidates' Spanish-language scaled scores are comparable to the scores of candidates who take the English-language GED tests. In 1995, about 26,500 Spanish-language GED tests were taken on the mainland United States and about 14,600 were taken in Puerto Rico. The paper describes the processes followed to analyze three forms of the five-test GED battery to determine if items are translatable, to ensure that all items are valid, and that the resulting instruments measure comparable constructs. The process of adapting the Writing Skills subtest is discussed in some detail. In addition, a linking design is outlined that introduces a procedure for screening biliterate students for equal proficiency in the two languages before including them in the linking sample. (Contains 2 tables and 16 references.) (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 414 320

Developing Parallel Tests Across Languages: Focus on the Translation and Adaptation Process

Joan E. Auchter
GED Testing Service
American Council on Education
1 Dupont Circle
Washington, DC 20036
ph 202-939-9400

and

Charles W. Stansfield
Second Language Testing, Inc.
10704 Mist Haven Terrace
N. Bethesda, MD 20852
ph. 301-231-6046

TM027838

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Requests for reprints may be sent to Joan E. Auchter, GED Testing Service, American Council on Education, 1 Dupont Circle, Washington DC, 20036.

BEST COPY AVAILABLE

Abstract

This article describes the GED Testing Service's Spanish Test Adaptation Project. The purpose of this project is to develop, based on an analysis of the issues, Spanish-language versions of the GED Tests that parallel the English language versions so that the GED candidates' Spanish-language scaled scores are comparable to the scores of candidates who take the English-language GED Tests. The article describes the process followed to analyze three forms of the five-test GED battery to determine if items are translatable, to ensure that all items are valid, and that the resulting instruments measure comparable constructs. The process of adapting the Writing Skills subtest is discussed in some detail. In addition, a linking design is outlined that introduces a procedure for screening biliterate students for equal proficiency in the two languages before including them in the linking sample.

Developing Parallel Tests Across Languages: Focus on the Translation and Adaptation Process[1]

The translation or adaptation of achievement tests into examinees' native languages is becoming more common as educators and testing organizations respond to the increasing diversity in the US population. The vast majority of articles on test translation and adaptation present the results of statistical comparisons and analyses of the dual language versions. Only a few publications on the test translation and adaptation process are available. Most of these focus on the translation of instruments other than achievement, such as measures of attitudes, intelligence, and personality traits. The constructs measured by such instruments can vary considerably across cultures. Achievement tests assess the mastery of a specified content domain. Therefore, such tests may be more amenable to direct translation, and the problems encountered in translating or adapting them may be of a somewhat different nature. This article describes the process followed by the GED Testing Service in translating and adapting the five GED Tests, which are tests of educational achievement.

BACKGROUND

Description of the English-language GED Tests

The GED Tests are designed to provide an opportunity for persons who have not graduated from high school to earn a high school level diploma that is recognized by both institutions of higher education and by employers. Administered in all fifty states and the territories in the United States, and in most Canadian provinces and territories, almost 800,000 people take the GED Tests annually. Approximately one of seven high school diplomas issued annually in the United States is a GED diploma.

The third generation of GED Tests, introduced in 1988, is a five-test battery that requires 7 hours and 45 minutes of test administration time. A GED candidate earns a GED diploma only after passing all five tests. The official titles of the five separate subject tests, and their time limits, are as follows: Test 1: Writing Skills (120 minutes), Test 2: Social Studies (85 minutes), Test 3: Science (95

---

[1]An earlier version of this paper presented at the 27th Annual Large Scale Assessment Conference, June 15-18, 1997, Colorado Springs, Colorado. Joan E. Auchter is Acting Director of the GEDTS. Charles W. Stansfield is President of SLTI is directing the Spanish language test adaptation project for the GEDTS.

minutes), Test 4: Interpreting Literature and the Arts (65 minutes), and Test 5: Mathematics (90 minutes). The Writing Skills Test has two parts: Part I is made up of multiple choice questions and Part II is a direct writing assessment (essay); the other four tests contain only multiple-choice questions.

To allow GED candidates the opportunity to demonstrate achievement comparable to that of high school graduates, the tests are based on two foundations: 1) test content that conforms as closely as possible to the core academic curricula of the U. S. high schools, and 2) score scales are based on periodic norming of the GED Tests on a nationally representative sample of graduating US high school seniors. This norming process allows the passing standards for the GED Tests to be referenced to the actual performance of those who graduate via the traditional route. The minimum passing score is set so that approximately 66% of graduating US high school seniors would pass the test battery and 34% would fail.

Description of the Spanish-language GED Tests

The Spanish-language GED Tests were originally developed to provide for adults in Puerto Rico, who had not graduated from high school, an opportunity to earn a GED diploma comparable to the diploma awarded by the high schools in Puerto Rico. (Spanish is the primary language of instruction in Puerto Rican high schools.) The 1988 revised Spanish-language GED Tests, introduced with the revised English-language tests, include content changes recommended by the Puerto Rican curricular experts and content specialists involved in development of the tests. For a direct comparison of the tests, see *The Tests of General Educational Development Technical Manual, First Edition,* (Auchter, 1993).

The Spanish-language GED Tests were normed using only graduating high school seniors in Puerto Rico. Due to the increasing number of Spanish speaking adults throughout the US without a high school diploma, many states began offering the Spanish-language GED Tests to their Spanish-speaking GED candidates. Currently, the Spanish-language GED Tests are taken more often in the continental US than in Puerto Rico. In 1995, about 26,500 Spanish GED Tests were administered in the mainland US and about 14,600 were administered in Puerto Rico (GEDTS, 1996).

STATEMENT OF THE PROBLEM

The Spanish-language GED Tests perform well for what they were developed and normed to do: provide an opportunity for adults in Puerto Rico to earn a GED diploma comparable to the diplomas awarded by high schools in Puerto Rico. The use of the Spanish-language tests outside of Puerto Rico

has been criticized because some states offer the same high school level credential, regardless of the particular language version of the GED Tests taken. This use is considered inappropriate because the content of the two test versions varies, the two language versions of the tests are normed on different populations, and the score scales are not linked. Thus, it is possible that different levels of ability are required to obtain the same GED score, and therefore, the credential. As a result of these concerns, both the GED Advisory Committee and the Commission on Educational Credit and Credentials, the governing board of the GED Testing Service, required the GED Testing Service to produce a new Spanish-language version of the GED Tests that will be comparable in content and difficulty to the English-language GED Tests.

METHODS

To determine if the goals of this project are obtainable, the GED Testing Service first conducted a preliminary analysis of the translatability of the GED Tests, and then commissioned three feasibility panels to explore the technical issues involved in linking the two language versions of the GED Test Battery.

Translatability Study

Prior to convening the three feasibility panels, GEDTS contracted with a translation firm to conduct a preliminary study evaluating the feasibility of doing a forward-translation of the English-language GED Tests into Spanish. The purpose of the study was to discern whether or not test items were amenable to translation to Spanish. In the study, all stimuli and each item and option in three forms of the GED were analyzed for translatability. The study concluded that the entire battery could be directly translated with minor modifications. In retrospect, it is clear that the findings of the study concerning Test One were somewhat optimistic. Complete analyses are included in the *GED Direct Translation Feasibility Study* (Colberg,1993).

Feasibility Panels

After it was clear that the language and content of the test could be translated, the GEDTS commissioned the formation of a series of feasibility panels to explore the technical issues involved. The first panel, the Psychometric Feasibility Panel (PFP), was convened in October 1993 to investigate

the feasibility of linking the English and Spanish-language versions of the GED Tests. The PFP consisted of the following four prominent psychometricians who were selected for consultation based on their experience in psychometrics and test equating: Linda L. Cook of the Educational Testing Service, Ronald K. Hambleton of the University of Massachusetts at Amherst, David Thissen of the University of North Carolina at Chapel Hill, and Howard Wainer of the Educational Testing Service. This panel met for two days, during which it discussed seven tentative linking options, evaluated IRT and anchor item linking procedures, and evaluated the utility of matching in linking and made a number of recommendations concerning linking options. The committee's deliberations are summarized in a document, *Linking the English-language and Spanish-language Versions of the Tests of General Educational Development: Psychometric Feasibility Study,* (Sireci, 1994).

A second panel, the Linguistic Feasibility Panel (LFP) which met in March 1994, considered translatability and other issues concerning the development of the new Spanish tests, and also made a number of recommendations. The LFP panel members, selected for consultation based on their expertise in linguistics and experience with language testing, included: Brunilda deLeon of the University of Massachusetts, Pardee Lowe, Jr. of the US Government Language School, Cecelia Rosenblum of the Educational Testing Service, Ramon Santiago of Lehman college, Charles Stansfield of Second Language Testing, Inc., and Gillian Whalen, a Spanish linguist and independent scholar. The LFP panel agreed with the translatability study's conclusion that most items can be translated. However, it differed somewhat in its conclusions. The LFP noted that some items in the Writing Skills Test would not test meaningful or challenging content in Spanish and that the validity of the translated Spanish test must be verified by professors of Spanish. The LFP outlined an appropriate translation process and made recommendations concerning the feasibility of translating the Writing Skills Test into Spanish. The panel noted that the Spanish-language version of the Writing Skills Test would tap different but related constructs; writing skills in English and writing skills in Spanish. While the constructs both relate to writing skills in the native language, they are nonetheless quite different abilities. Thus, LFP advised that it may not be possible to report the score on the Writing Skills Test in the same way it is to report the score on a translated test of a subject other than language. Clearly, the development of a Spanish language version of the Writing Skills Test would involve adaptation; in other words, a significant modification of the instrument. The LFP deliberations are summarized in the *Development of Revised Spanish-language Versions of the Tests of General Educational Development: Linguistic Feasibility Study* (Auchter, 1994).

Subsequently, it was decided to convene a third panel to consider the findings of both previous panels, and to make more detailed recommendations as to how to proceed with the

translation and linking of the two language versions. The Combined Feasibility Panel (CFP), met for two days in October 1995. The following CFP members were selected based on their expertise in linguistics, second language testing, cross-lingual assessment, and psychometric methods: Ronald Hambleton of the University of Massachusetts, Pardee Lowe, Jr. of the US Government Language School, Maria Pennoch-Roman and Alicia Perez Schmitt of the Educational Testing Service. The CFP set desirable background characteristics of translators, and reviewed sampling issues, linking designs, and procedures to check for comparable standards. The CFP agreed on an approach to the translation and linking based on the LFP report and the *Guidelines for Adapting Educational and Psychological Tests* from the International Test Commission (in press; summarized by Hambleton, 1994). This approach is reported in the *Options for the Development and Linking of New Spanish-Language Versions of the GED Tests* (Auchter and Stansfield, 1996). Subsequently, the CFP reviewed and approved the *Action Plan for the Development and Linking of New Spanish-Language Versions of the GED Tests* (GEDTS, 1996) that is discussed later. While the CFP's purview was the entire test, approximately half the panel's deliberations focused on GED Writing Skills Test.

TRANSLATION PROCESS

The translation of a test into another language is an important task. It is assumed by test score users that the translated items are equivalent in meaning and difficulty to the original version in English. This equivalence reinforces the claim for score comparability. If the translation is accurate, then the examinee will not be affected (assisted or disadvantaged) by the quality of the translation. Thus, the examinee's response to each item will reflect the ability to respond in his or her native language to the exact same item that was administered in English to English proficient students.

Similarly, a translation must be expressed in natural language, or in language that is as natural as the language used in the English original. If a translation is too literal, it will read like a translation as opposed to an authentic document in the target language. This lack of naturalness in the wording of the item often results in poor quality items which, generally, are more difficult. Haladyna (1994, p. 64) points out that unedited, awkwardly written items tend to distract some test takers by causing them to loose their concentration. Haladyna states: "This inattention produces a bias in test scores that undermines the valid interpretation or use of test scores." Furthermore, research on item bias on the NTE has shown that it is often the least able examinee who is most disadvantaged by awkwardly worded items (Wolfram, Figueroa & Christian, 1991).

The same concerns are relevant to test translation. If a translation is too literal, then the

meaning of the original item will be distorted because a critical distinction in the original may be simplified or not carried over to the translated tests. Normally, a distortion in meaning makes it less probable that the examinee will perform well on the item. The resulting loss of information usually makes the item harder to answer correctly. Sometimes a translated document may be more clear, because of efforts to improve its meaningfulness. This can actually result in easier items (Stansfield, 1996). Sireci (1997) observes that increased rigor in the translation process may facilitate item equivalence across languages. For additional information on the translation of tests, see (Hambleton, 1993,1994).

Our presentation of the translation process is divided into two broad groupings: general guidelines which apply to all five subtests of the GED, and issues that are specific to individual subject-area subtests. The general guidelines are addressed first and the issues specific to particular subject-area subtests follow.

General Guidelines in Translation

While there are issues specific to each of the five subject-area tests, the following steps applied to all five tests.

Step One: Selecting Three Forms Most Appropriate for Translation. The two reviewers were selected to evaluate existing forms of the GED to determine those forms most suitable for translation or adaptation to Spanish. The reviewers were native Spanish-speakers who have worked on the GED Tests for a number of years. In addition to a native command of Spanish, both reviewers have extensive test development experience and experience as professional translators. Seven equated operational forms of the Science, Social Studies, Mathematics, Interpreting Literature and the Arts, and Writing Skills Tests were reviewed. The criteria used to review the forms were 1) the degree to which the tests reflect recent changes in the test specifications, 2) the accessibility of test content to Hispanic examinees, and 3) the ease with which the language of the test could be rendered into Spanish. Based on the review, the three most suitable forms of each subtest were selected for the project.

Step Two: Selecting Translators. After the test forms were identified, the principal translators were selected. The requirements for the selection of translators, as defined by the LFP and CFP, included 1) American Translation Association (ATA) or equivalent certification, 2) near-native

reading and writing skills in English, the source or donor language, 3) educated native writing skills in Spanish, the target or receptor language, and 4) congruity judgment, which is the ability to judge the equivalence of the original and translated text in terms of their meaning, style, rhetorical structure, and linguistic complexity, 5) experience in the test development process - ideally, experience as an item writer; translators should be familiar with the mechanics and rules of item writing, including the role of grammatical clues in the wording of items, clang associations,[2] the length of the correct answer, and the homogeneity and parallelism of the options, 6) appropriate academic training and subject specialization; different translators were selected according to their area of specialization.

Step Three: Translator Orientation. Because tests represent a different kind of text than translators routinely handle, the proper and detailed orientation of translators is especially important. Prior to beginning the translation, translators were given basic information on the GED Testing Program and the test population. Translators also were given a copy of the test specifications for the tests they were translating, the *Technical Manual for the Tests of General Educational Development*, the *Item Writer's Manual* furnished to English-language GEDTS item writers, and the *Guidelines for Adapting Educational and Psychological Tests* (Hambleton, 1994).

Translators were given a copy of the English-language versions of the tests, including graphics, and were requested to provide the translation of all text, including titles and footers. The importance of translating each message or proposition within each test stimulus or task was emphasized. Translations from English into Spanish all too frequently retain the use of the passive voice when it would be more germane to a Spanish-language text to use the active voice. The result is an anglicized text that is structurally inappropriate, and, hence, more difficult to read and comprehend.

Translators also were coached to be aware of dialect and syntax issues. Since GED examinees are expected from all Spanish-speaking countries, the translators were advised to make a conscientious effort to use language that is not biased toward the peculiarities of any particular national speech. The language should be as clear to a person of Argentine roots as to one of Mexican or Spanish heritage. Terms that vary across dialects also pose a considerable problem that translators must address. In this case, it was decided to consider all possible variants of a word or phrase, and then to look for the variant that is most neutral or most widely understood across the

---

[2]Clang associations is a term that refers to the construction of distractors based on the repetition of words that occur in the stimulus.

Spanish-speaking world. An example is the word for car in Spanish. Depending on dialect, a speaker might commonly use *coche, carro, or maquina*. Each of these words could mean something different to speakers of the other dialect. Yet the word *automovil* would be understood by speakers of all these dialects.

Step Four: Initial Forward-Translation. When translating a test, a testing company is faced with the issue of whether to contract for a back translation as a quality control mechanism. Back translation is sometimes used in the development of foreign language versions of tests and questionnaires. The literature on it comes not from the field of translation studies, but from cross-cultural psychology. Brislin (1986) has written extensively on back translation and Hambleton (1994) speaks favorably of it. Forward translation involves rendering a source document into the target language.[3] Back translation involves rendering the forward translation back into the source language. The back translation is then compared with the source document in order to identify discrepancies between the two. The forward translation is then examined to see if it is the cause of each discrepancy. When the forward translation is determined to be the cause of the discrepancy, the forward translation is revised.

After discussing the issue of back translation, the LFP and PFP agreed that forward-translation followed by successive stages of review and revision was the most appropriate procedure to follow. Each form of each subject-area test was translated from the source to the target language by the principal translator. The primary translators were also requested to compile a file of comments identifying any items in the tests that could not be translated, as well as any items or portions of the tests that posed special problems for translation, and how these were handled.

Step Five: Initial Review. This initial translation and the translator's file of comments were reviewed by a primary reviewer who was asked to judge the congruity of the translation with the English-language version of the test. Again, each reviewer was a specialist in the translation of the

---

[3]In translation studies it is common to speak of the source language or document and the target language or document. The source language is the language of the original document; the target language is the language into which the translator renders the document. The verb "render" is used to mean "translate" in the translation literature. However, its usage implies that the task is not a process of word-for-word translation. Rather, "rendering" involves creating a document with equivalent meaning and style appropriate to the target language. Rendering implies a document that does not read like a translation.

subject area of the test. The reviewer was tasked to create a list of specific concerns and suggested revisions. This list was then returned to the project manager.

Step Six: Translation Contractor Review. Each test was returned to one of the two translation contractors who have extensive experience in test development and translation. After reading the translation, the translator's comments, and review, each contractor's project manager discussed the issues with each reviewer and then with the primary translator. The primary translator revised the translation based on the suggested revisions. The primary translator kept a file on each suggested revision. The file indicates whether the revision was implemented or not. If the revision was not implemented a justification for this decision was provided.

Step Seven: Secondary Review. The revised translation was then reviewed by two secondary reviewers contracted by the GEDTS. These reviewers were selected because of their familiarity with the subject and their sensitivity to Spanish dialects and to regional and cultural differences. These reviewers reviewed the translations for linguistic accessibility, equivalence of meaning, and naturalness of expression in Spanish. They described problems in a memorandum and suggested revisions where appropriate on the manuscript. These problems and suggested revisions were returned to each project manager, who, after reviewing them, returned them to the primary translator. Again, the primary translator either made the revisions or documented why the revisions were not made.

Step Eight: Key Verification. With the translation in final form, the primary translator and one reviewer read each test and marked the correct response for each test item. Then, the two keys were compared with each other and with the English original. This step added additional verification to the translation and corroborated the viability of all correct answers and distractors, thus corroborating the preservation of the original and the integrity of the instrument.

Step Nine: Translation Documentation. Because the quality of a translation is critical to the reliability, validity, and score comparability of a test, it is necessary to document the process that was followed to translate a test. Each translation contractor was required to document the process and efforts that were made to ensure the quality of the translation of each form of each of the GED subject area tests, as well as the professional qualifications of the translators who performed the translations and the reviews. This documentation took the form of a formal report to the GEDTS by

each contractor.

Issues Specific to Individual Subject-Area Subtests

In addition to the general steps described above, issues are relevant to the description of the translated versions of the specific subtests are described below. Because of it's unique nature, the adaptation process for the Writing Skills Test is described last.

Test 2: Social Studies. While the Social Studies Test includes fewer technical terms than the Science Test, its length and content pose significant translation challenges. Since much of the content is related to US history and culture, it can pose challenges for rendering to another language. For example, terms such as "freedom rider" have no equivalent in Spanish, resulting in the need to paraphrase or define. Each primary translator made a detailed list of such words or concepts and how they were handled in the translation.

There were no passages that could not be translated and no items in any of the three translated forms that required major changes or substitutions. Therefore, all passages and items on the Spanish-language Social Studies test are direct translations and can potentially be considered as anchor items.

Test 3: Science. A viable, faithful translation of a science test depends on the translator having a strong science background and knowledge of how science concepts are expressed in the target language because of the technical terms included in the source text. However, there are many classifications in the taxonomies of the animal and plant kingdom that escape the memory of the strictest specialists (Colberg, 1996). Therefore, the science translator must have access to reference materials and specialized glossaries. The science translator must also have an adequate command of scientific usage to select the less technical but more commonly used term to express a scientific concept or phenomenon when appropriate to the test and the examinee population. The translators used for Test 3 were specialists in scientific and technical translations, while the reviewers were specialists in the bilingual teaching of science at the secondary level. Because the expression of scientific phenomena is formulaic in nature, there were few disagreements between translators and reviewers. The major contention was over the term for carbon dioxide, which is expressed with two levels of formality in Spanish. Again, the less formal level was selected for use in the test, followed by inclusion of the more formal term in parentheses as a gloss.

There were no items that could not be translated and no major changes in items or item substitutions on the three Science Test forms. Therefore, all items on the Spanish-language Science Test are direct translations of the English-language items and can be considered as potential anchor items.

Test Four: Interpreting Literature and the Arts. Like test translation, literary translation requires special skills. The Interpreting Literature and the Arts test required a literary translator, someone who is also skilled as a creative writer in the target language. The method of translation followed with Test Four was essentially the same as the method described above, although the translation of unpublished poems and dramatic excepts required greater interaction among translators and reviewers.

Prior to the translation of the literary selections, an extensive search was conducted to ascertain whether or not a published translation was available. If more than one translation was available (as was the case with works by major literary figures), these translations were evaluated to determine which was best by analyzing and comparing the translation with the English-language original in order to ensure that the translator had been totally faithful to the original text and that the style and complexity of the two versions were congruent. Those selections for which published translations were not found were translated with care to preserve the meaning of the text as well as its aesthetic qualities.

In the translation of the Interpreting Literature and the Arts Test, no texts or items were identified by the translator as unsuitable for the translated version. Therefore, all items on all three forms are direct translations of the English-language items and can be considered as potential anchor items.

Test Five: Mathematics. The different forms of the mathematics tests were translated by professional translators with degrees in mathematics from a Latin American university.

There were no translation issues with the Mathematics Test. The most contentious issue to arise was how to translate "right triangle." Since two different terms are used in different countries of the Hispanic world without overlapping usage between countries, the translation contractors decided to use one term, and to insert the other in parentheses as a gloss. Another source of contention between translators and reviewers of the different forms was the Spanish word for "rate of interest." Again, different levels of formality exist, with one being used by the general public and another being used by accountants and economists. The term used by the general public was

ultimately selected.

WRITING SKILLS TEST ADAPTATION PROCESS

Description of the English Version

Because it was necessary to modify Test 1, Writing Skills, it is appropriate to begin with a detailed description of the format of the English version of GED Test 1.

The Writing Skills Test is a two-part test. The multiple-choice portion consists of 55 items that test knowledge of the structure and conventions of standard written English. The essay portion consists of a 45 minute writing sample based on a specific prompt. A dozen English prompts that functioned well during pretesting were reviewed for their accessibility to a Hispanic examinee population. About half the prompts were judged to be about equally accessible to Hispanics. These prompts were translated to Spanish and are used with the current Spanish-language versions of the Writing Skills Test.

Examinee writing samples are rated using a six point holistic scoring guide. The guide was previously translated to Spanish for use with the current Spanish language GED Tests developed for Puerto Rico. Since it was first translated, it has been used many times by the GEDTS pool of bilingual readers, and over the years, several minor revisions and additions have been made to better reflect the linguistic features of standard written Spanish. The prompts and the scoring guide are usable as part of the new Spanish language version of Test 1.

Multiple-choice items in Test 1 items are based on a stimulus text which contains various kinds of errors in writing that have been deliberately introduced into the text. For each sentence containing an error, the examinee must choose from five options the one that would make the sentence correct. Many of the translated items are valid measures of knowledge of the structure and conventions of formal written Spanish. For example, items that test subject-verb agreement or coordination of tenses across clauses in English will normally test similarly valid knowledge and skills when translated to Spanish.

However, other items pose a problem when translated. Items that test spelling in English are often less valid, because the particular word that contains the spelling problem does not pose a spelling problem when translated into Spanish. An example is knowledge of the difference between "there" and "their." If one of these were tested and the other used as the basis for distractors on the Spanish exam, probably no examinee would confuse the usage of the three Spanish equivalents,

"hay," "su," and "allí." Similarly, for the distinction between the usage of "would" and "wood," examinees would never confuse "madera" with "hubiera." So it was clear from the results of the translatability study that some translated items would have to be modified or even completely replaced. Finally, written Spanish contains diacritic marks not found in written English. These are just a few of the ways in which Spanish differs from English.

Initial translation

In order to deal with these issues, the GEDTS initiated an iterative test adaptation process, which has recently been completed. The first step was the selection of an appropriate translator. The International Test Commission's *Guidelines for Adapting Educational and Psychological Tests* (Hambleton 1994) call for the use of translators who know both languages and cultures, the content of the discipline in which the translation will be done, and experience in item writing and the test development process. The translator selected to translate Test 1 was a native speaker of standard Latin American Spanish who has lived in the US for 25 years, an experienced professional translator accredited by the American Translators Association, has taught Spanish at two universities before becoming a professional translator and interpreter, and has experience as an item writer and developer of tests for the selection of translators by US Government agencies.

Once the lead translator had been identified, the next step was the selection of the forms of the test to be translated. First, GEDTS staff selected seven forms of the test that reflected the latest revisions to the test specifications and that had shown good psychometric characteristics through pretesting and operational administration. These seven forms were then reviewed by the translator who did the translatability study mentioned earlier. She was tasked with identifying the three forms most suitable in content and language for translation. These forms were then translated to Spanish by the primary translator selected for this project. The primary translator was instructed to translate all items whose content was translatable and to modify items or options that were not translatable using the same stimulus sentences that were used in the English original.

Upon completion of this work, the translation was reviewed by another experienced translator who served as a primary reviewer. Suggested corrections and revisions were either implemented or a written explanation as to why they were not implemented was provided in a separate document. In addition, the primary translator created a document that discussed any difficulties surrounding the translation of each item and option. The document also classified the translated item according to the content specifications for GED Test 1 and each item was classified

BEST COPY AVAILABLE

as being the same as the English original or a modification.

Role of the Test Advisory Committee

The GED Testing Service then convened a four person national Test Advisory Committee for the Spanish language version of Test 1. The role of the committee was to assist in the adaptation of Test 1 by specifying the degree of adaptation needed through recommendations for an adapted set of specifications, and to approve all items that would appear on the test. All members had extensive experience teaching Spanish to native speakers in this country at either the high school, community college, or college level and all were experienced in the test development process. Three of the four had previously served as full-time employees of another test publisher and three of the four were native speakers of Spanish.

The Advisory Committee members were sent background information on the GED program, background information on the deliberations of the three panels that had been previously convened to discuss the translation/adaptation of the GED tests, the English original and translations of each form of Test 1, a content analysis for each form showing how each item is classified in the specifications and its p value, and the *Test 1 Item Writer's Manual* in English which contains the test specifications. They were instructed to review the *Item Writer's Manual*, the English language test forms, and the Spanish language translations, and to make detailed notes and comments concerning revisions that should be made in the translated items. They were also instructed to identify any items that would be inappropriate for a Hispanic examinee population.

Recommendations of the Committee

The Advisory Committee met for a total of 18 hours during two days in October 1996. The first day of the meeting was devoted to reaching consensus on the wording of each stimulus text and each item on one form of the test. The discussions were based on the detailed comments that committee members had written on the tests they had been sent. Members also classified each item as being either a direct translation from English, a minor change from English (involving only one option), or a major change from English (involving two or more options). For the one form examined, it was found that 67% of the items initially provided by the translator/item writer were direct translations of the English original, 22% involved a minor change, and 11% a major change from the English original.

The second day of the meeting was devoted to a discussion of how well the

English-language test specifications applied to the Spanish language version. The English-language test specifications are presented, along with examples, in the Item Writer's Manual, which was used as a point of departure. Committee members began by reviewing the three item types on the Writing Skills Test. They pointed out that often one item type seemed awkward in Spanish and suggested modifications to alleviate the problem.

Next, they discussed the content categories included in the specifications for the English language version. As they progressed through each category, they indicated whether it applied to Spanish and to the same degree, i.e., whether a content category was more or less important in Spanish than in English. They identified categories that could be deleted from the Spanish version of the Manual and added several new categories. Some of the new categories are a) Gender and Number Agreement of all types, b) Prepositions, c) Other Troublesome Words, and d) Accents and other diacritic marks.[4]

The review of content categories produced a shift in the general content distribution, which is depicted in Table 1. The shift in percentage of items devoted to each content category involved an increase in the number of items devoted to Sentence Structure. Items testing Sentence Structure focus on grammatical constraints that operate across the clause level. Such items would test the correct use of coordination and subordination in and across sentences.

INSERT TABLE 1 ABOUT HERE

Another important issue dealt with by the committee was the kind of examinee for whom the item is being written. Examinees taking this test may be either monolingual or bilingual. While the two groups show considerable overlap in the kinds of errors they make in writing, bilinguals make some errors that are unique only to bilinguals. It was decided to write items that would test common confusions of the monolingual only. Otherwise, bilinguals would be disadvantaged, since they would find both types of items challenging, while monolinguals would only find one type of item challenging.

Finally, the committee discussed the issue of different oral Spanish dialects. It was

---

[4]Gender and Number Agreement would involve the testing of predicate adjectives (Ella está cansada.), all cases of pronouns, including interrogative and neuter forms, related forms of adjectives (demonstrative, possessive), etc. Preposition include phrasal verbs and the use or prepositions with interrogative and relative pronouns. Other Troublesome Words involve the use of words whose gender is irregular (mapa, clima, cura, agua) or may change in different contexts. Accents was placed under the Category of Spelling, rather than Punctuation. The same applies to the use of the tilde and the dieresis. Other diacritics were placed under the Punctuation category. These include inverted question and interrogation marks.

decided that the testing of nonstandard but widely used verb forms such as "haiga" for "haya" was appropriate. On the other hand, it was decided that the spelling problems tested should not be dialect based, since this would disadvantage the speakers of those dialects. For example, only Puerto Ricans might confuse l and r because of their substitution in Puerto Rican spoken dialect. Because Spanish has some standard dialects with unique features, use of the forms of those dialects should not be tested, since that would advantage speakers of those dialects. Thus, no verb forms associated with vos, an alternative form of the second person singular used in Central America and southern South American, should be tested; nor should be vosotros forms used in Spain. On the other hand, errors associated with the language development process (child-like speech), such as "cabo" for "quepo," should be tested.

Other Related Work

Detailed minutes of the Advisory Committee meeting were prepared and sent to the committee for review, revision, and eventual approval. Subsequently, the minutes and copies of the detailed comments on each item were sent to the translator/item writer. These were used to revise the tests and to create new items to fit the new content categories. The primary translator then prepared a revision of the Item Writer's Manual for the Spanish Test. This was titled the *Item Writer's/Translators Manual for the Spanish Language Version of Test 1*. It includes the revised content specifications, new or modified content categories, and examples in Spanish of stems for each item type within each content category. It also establishes the exact wording of formulaic expressions that appear repeatedly in item stems and options.[5] It includes a sample translation of a text, the associated sample items and discusses how to use existing text to test new content categories.

The Advisory Committee was very pleased with the additional revisions, the new items written, and with the Item Writer's/Translator's Manual. Naturally, however, members still had some additional revisions to suggest in the wording of the translated stimulus texts, and they identified a few items which they felt might have double keys. Their concerns have been accommodated, and the test is now finalized.

---

[5]Examples of these would be the translated versions of a.) remove the comma after..., b.) insert a comma after..., c.) change .,. to ... d.) replace ... with ... e.) change the spelling of ... to ... f.) no correction is necessary. Formulaic wording also includes the translation of leads like g.) What correction should be made to this sentence? h.) Which is the best way to write the underlined portion of this sentence I.) If you rewrote sentence..beginning with ..., the next word(s) should be... It is important that the wording of these be standardized across different forms of the test.

ANALYSIS OF CHANGES IN ITEMS

The final version of the test was then analyzed to determine the relationship of each item on the translated version to the original English version. This content analysis shows for each item, the item type used and the content category and subcategory tested. It also gives an analysis of the similarity of the translated version to the English original for each item and option.

Table 2 below depicts the results of the analysis for the total test for the three forms. Items are classified as

- direct translations from English or same as English (SAE)–judged to be content valid in both languages,
- exhibiting a minor change (MC) involving only one distractor,
- exhibiting a big change (BC) involving changes on two or more distractors,
- testing content categories that are unique to the Spanish specifications (S), through use of the translated version of the original English stimulus sentence.

Cases where new stimulus sentences were used to construct new items are also identified, and they are subdivided by those that test Spanish specific content categories (New-S), and those that test content categories that also appear on the English test (New). Sometimes items of the latter type were constructed

- to ensure content balance within or across the forms,
- because the wording of a translated item was awkward,
- because the translated item would clearly have been easier or more difficult in Spanish, or
- because the item subtype itself posed unique problems when converted to Spanish.

INSERT TABLE 2 ABOUT HERE.

The results of the comparison of the English and Spanish versions across the three forms shows that almost half of the original items were directly translatable to Spanish. About 20% of the items required that two or more distractors be changed, and another 20% of the translated stems resulted in changes that reflect Spanish only content categories. The remaining 10% of the items involve new stimulus sentences and options. One item required only a change of one distractor.

ESTABLISHING EMPIRICAL LINKS

During the winter of 1998, the tests will be administered to a large sample of biliterate Hispanic high school seniors in Florida, California, New York, Texas, and other states with a large Hispanic population. Subjects in the sample will be selected based on the similarity of their performance on a screening test in both languages. The screening test will be Test 4 of the GED, which is a measure of reading comprehension involving passages from literature and the arts. Each examinee will receive a bilingual test booklet, with half of the items in English and half in Spanish. The order of presentation will be counterbalanced. Examinees whose score on each half differs by only two standard errors of measurement (3 raw score points) will be considered to be balanced biliterates. From this group, a sample will be selected which is as similar as possible to the distribution of ability within the sample of 12th grade students that was used to establish norms and the cut score for the GED in English in 1996. These students will then take different forms of Test 1 in both English and Spanish. The order of presentation will again be counterbalanced. Test 1 English item parameters from the biliterate sample will be will linearly transformed to the same scale as the 1996 standardization. Items on the Spanish version that have been identified as identical to those on the English version, will be considered for use as an anchor test. The tentative anchor items on the Spanish test will be linearly transformed to the 1996 standardization scale. Those with similar item parameters will then serve as an anchor to calibrate the scores on the Spanish language version and link it to the score scale for the 1996 standardization of the English language version. In this way, we expect to be able to assure score users that scores on the Spanish test reflect a degree of mastery of the construct that is comparable, within a specified error of measurement, to the equivalent scores on the English language version of Test 1. The use of a biliterate sample will permit us to examine whether the care that has gone into the translation process has maintained comparable psychometric charactistics for items that were not modified in Test 1 and in the other four GED Tests. We hope to report on these matters later.

CONCLUSIONS

After working on the Spanish language GED test adaptation process for three years, we can make the following observations that may be of assistance to others involved in test translation projects:

- Begin with a professional analysis of the translatability of each section of the test and all items in each section. Document the findings of the analysis.
- Select professional translators who are specialists in the subject of the test and

experienced in test development and test translation.

- Develop guidelines and procedures that require a rigorous translation from the source language to the target language. The procedures should require extensive review and revision in an iterative process. Then, follow the guidelines strictly. In that way, differences in item difficulty introduced when translating tests from one language to another are minimized.
- Document all revisions to test stimuli, stems, and options for reference in future test and item analyses.
- Document the entire process in a final report. This report will contribute to the evidence or any claims made concerning score comparability.

Test translation is generally done long after the source language test is developed and standards are set. Our experience in this project suggests that the following guidelines for future English language test development that would smooth the way for translation of new GED test forms to other languages:

- Avoid stimuli that reference topics identified with American culture, such as baseball. For example, in baseball, which is not an international sport, there is no translation for "shortstop." A careful review of all stimulus should be done prior to developing items.
- When possible, select literary texts for which a translation already exists. By using texts that have published translations, the time spent translating would be eliminated.
- Add translation reviewers to the item and test review stages of test development. These reviewers can identify potential translation problems and suggest revisions during the test development stages.

While translation from one language to another does not automatically result in tests that are equivalent in both languages, careful attention to translation issues during the English language test development process and strict adherence to established translation guidelines can reduce the likelihood of introducing factors that can lead to differences in test performance, validity, and score comparability.

References

Auchter, J.C. & Stansfield, C.W. (1996). *Action plan for the development and linking of new Spanish-language versions of the GED Tests.* Washington, DC: GED Testing Service.

Auchter, J.C. & Stansfield, C.W. (1996). *Development of revised Spanish-language versions of the Tests of General Educational Development: Linguistic feasibility study* Washington, DC: GED Testing Service.

Baldwin, J. (1996). *Who took the GED?* Washington, DC: American Council on Education.

Brislin, R.W. (1986). The wording and translation of research instruments. In W.J. Loner & J.W. Berry (Eds.), *Field methods in cross-cultural research (pp. 137-164).* Newbury Park, CA: Sage.

Colberg, M. (1993). *Direct translation feasibility study.* Washington, DC: GED Testing Service.

Colberg, M. (1993). *Translation of the Tests for General Educational Development: Documentation Report by Logos, Inc.* Washington, DC: Logos, Inc.

GED Testing Service. (1990). *Item writer's manual, Test 1: Writing Skills.* Washington, DC: Author.

GED Testing Service. (1997). *Item writer/translator's manual for the Spanish language version of Test 1: Written Expression.* Washington, DC: Author.

Haladyna, T.M. (1994). *Developing and validating multiple-choice test items.* Hillsdale, NJ: Lawrence Earbaum Associates.

Hambleton, R. (1993). Translating achievement tests for use in cross-national studies. *European Journal of Psychological Assessment*, 9(1), 57-68.

Hambleton, R. (1994). Guidelines for adapting educational and psychological tests: A progress report. *European Journal of Psychological Assessment*, 10(3), 229-244.

Sireci, S.G. (1994). *Linking the English-language and Spanish-language versions of the Tests of General Educational Development: Recommendations of the GED-STEP Psychometric Feasibility Panel.* Washington, DC: GED Testing Service.

Sireci, S.G. (1997). Problems and issues in linking assessments across languages. *Educational Measurement: Issues and Practices.* 16(1), 12-19.

Stansfield, C.W. & Auchter, J.C. (1996). *Options for the development and linking of new Spanish-language versions of the GED tests: A report emanating from the deliberations of the Combined Feasibility Panel.* Washington, DC: GED Testing Service.

Stansfield, C.W. (1996). *Report to GEDTS on translation procedures employed by Second Language Testing, Inc.* N. Bethesda, MD: Second Language Testing, Inc.

Wolfram, W., Figueroa, E., & Christian, D. (1991). *Investigative research on sociolinguistic dimensions of the NTE.* Princeton, NJ: Educational Testing Service.

TABLE 1

Writing Skills Test Content Categories

Content Category	English	Spanish
Sentence Structure	25%	30%
Usage	45%	40%
Mechanics	30%	30%

TABLE 2

Classification of Writing Skills Items in Adapted Spanish Form

	FORM AM					
	SAE	MC	BC	S	New-S	New
Number Items	26		13	9	6	1
Percent	47%		24%	16%	11%	2
	FORM AQ					
	SAE	MC	BC	S	New-S	New
Number Items	26	1	11	10	5	1
Percent	47%	2%	20%	18%	9%	2%
	FORM AS					
	SAE	MC	BC	S	New-S	New
Number Items	27		11	13	3	1
Percent	49%		20%	24%	5%	2